Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Кафедра «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ

по лабораторная работа №3

по курсу «Математическая статистика»

Вариант №12

Выполнил студент: Мхитарян В.К.

Группа: ИУ7-64

Проверил: Велищанский М.А.

2019 г.

1. Постановка задачи аппрокисмации неизвестной зависимости по результатам наблюдений

Пусть Y – случайная величина, $X_1, \dots, X_s$– детерминированные величины.
Если изменение значений $X_1, \dots, X_s$ влияет на значения случайной величины Y, то говорят, что Y стохастически зависит от $X_1, \dots, X_s$. Задача регрессионного анализа – задача, связанная с установлением аналитических зависимостей между случайной величиной Y и детерминированными величинами $X_1, \dots, X_s$, носящими количественный характер. В регрессионном анализе используется модель черного ящика, как наиболее общая модель, ассоциируемая с понятием отображения. На вход поступает вектор $(X_1, \dots, X_s)$, который посредством некоторого отображения Ф и случайных возмущений $(\varepsilon_1, \dots, \varepsilon_s)$ преобразуется в вектор $(Y_1, \dots, Y_s)$.

2. Понятие МНК-оценки параметров линейной модели

Рассмотрим частный случай: m=1, s=1, имеются результаты n наблюдений;

$\begin{cases} y_1 = \Phi(x_1) + \varepsilon_1 \\ \cdots \\ y_s = \Phi(x_s) + \varepsilon_s \end{cases}$, где y1, $\dots, y_n$ — n реализаций Y (отклики); $\varepsilon_1, \dots, \varepsilon_n$— n реализаций ε (случайная величина, характеризующая случайные ошибки); $x_1, \dots, x_n$ – известные значения (факторы); Ф – некоторое неслучайное отображение

Часто в качестве функции $\widehat{\Phi}(x)$ выбирают функцию следующего вида:

$\widehat{\Phi}(x) = \theta_1\psi_1(x) + \ldots + \theta_p\psi_p(x),$ **где**

- $\psi_1, \ldots, \psi_p$ – известные базисные функции.

Параметры $\theta_1, \dots, \theta_s$ подбирают так, чтобы $\widehat{\Phi}(x)$ наилучшим образом аппроксимировала $\Phi(x)$. В этом случае регрессионная модель называется линейной по параметрам.

С учётом предположения о виде функции $\widehat{\Phi}$ результаты наблюдений можно записать в виде:

$y_i = \theta_1\psi_1(x_i) + \ldots + \theta_p\psi_p(x_i) + \varepsilon_i, \quad i = \overline{1;n}$.

В матричном виде:

$\overrightarrow{y} = \Psi\overrightarrow{\theta} + \overrightarrow{\varepsilon},$ **где**

$$\overrightarrow{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \Psi = \begin{pmatrix} \psi_1(x_1)\psi_2(x_1) & \cdots\psi_p(x_1) \\ \psi_1(x_2)\psi_2(x_2) & \cdots\psi_p(x_2) \\ \vdots\ \vdots & \ddots\ \vdots \\ \psi_1(x_n)\psi_2(x_n) & \cdots\psi_p(x_n) \end{pmatrix}, \quad \overrightarrow{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \overrightarrow{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Задача заключается в подборе $\overrightarrow{\theta}$.

Будем предполагать, что:

1. $M\varepsilon = 0$, т.е. систематические ошибки отсутствуют;
2. $\varepsilon \sim N(0,\sigma^2)$.

Оценка $\hat{\overrightarrow{\theta}}$ вектора $\overrightarrow{\theta}$ называется оценкой, полученной по методу наименьших квадратов (МНК-оценкой), если $\overrightarrow{\theta}$ доставляет минимальное значение функции S($\overrightarrow{\theta}$) = $\|y - \Psi\overrightarrow{\theta}\|^2$.

3. Формулы для вычисления МНК-оценки в рассматриваемом случае

В рассматриваемом случае МНК-оценка вектора $\overrightarrow{\theta}$ имеет вид:

$\hat{\overrightarrow{\theta}} = (\Psi^T\Psi)^{-1} \cdot \Psi^T \overrightarrow{y}$

так как $y = \theta_0 + \theta_1 t + \theta_2 t^2$, то

$$\Psi = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}.$$

Среднеквадратичное отклонение полученной модели от результатов наблюдений будем вычислять как

$\Delta = \sqrt{\sum_{i=1}^{n} \left(y_i - y(t_i)\right)^2}$, где

y_i – экспериментально полученное значение Y при $t = t_i$;

$y(t_i)$ – значение Y, полученное с помощью МНК-оценки.

4. Текст программы

```
function Lab31()
%% аппроксимация неизвестной зависимости параболой
   close all;
   T = importdata('t.txt');
   Y = importdata('y.txt');
   One(1:length(T), 1) = 1;
   T2 = T.^2;
   F = horzcat(One, T, T2);

   Ft = transpose(F);
   theta = Ft * F \ Ft*Y;

   Yt = theta(1) + theta(2) * T + theta(3) * T2;
   delta = sqrt(sum((Y - Yt).^2));
   deltaS = sprintf('\\Delta  = %.5f\n', delta);

   figure(1);
   %переопределим Yt, чтобы не получать кусочную функцию на малых выборках
   T_G = min(T):0.01:max(T);
   T_G2 = T_G.^2;
   Yt = theta(1) + theta(2) * T_G + theta(3) * T_G2;

   plot(T, Y, '.r'); %экспериментальные данные
   hold on;

   plot(T_G, Yt, 'b'); %полученная аппроксимация
   grid on;

   text(140,20, deltaS, 'Units', 'pixels');
   y_eq = sprintf('y = %.2f + %.2f*t + %.2f*t^2', theta(1), theta(2), theta(3));
   legend('Y experimental', y_eq);
   legend('Y experimental', y_eq);
end
```

5. Результаты расчетов и график

МНК-оценка в данном варианте имеет вид

$$\hat{\vec{\theta}} = \begin{pmatrix} 4.80 \\ 2.34 \\ 1.86 \end{pmatrix}.$$

В свою очередь среднеквадратичного отклонение полученной модели от результатов наблюдений равна

$$\Delta = 125.44$$

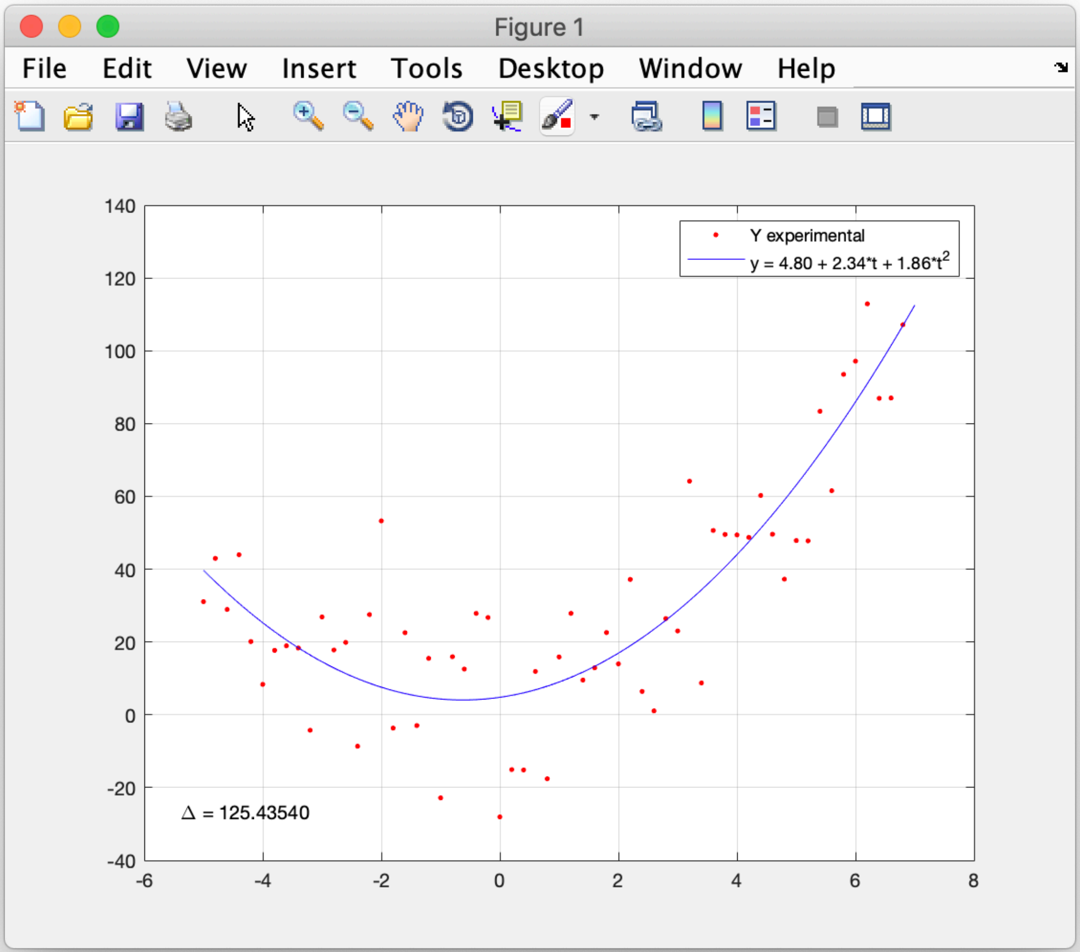

Рис.1 – График исходной выборки и полученной модели